SSOV and SSOF User's Guide

1. Description

SSOV and SSOF are FORTRAN77 programs that **compute analytically the expected adverse impact ratio, the expected quality and the standard deviation of the average criterion score of the selected in case of a simple multi-stage selection from a small heterogeneous applicant pool.** For details on the computation we refer to De Corte (2014b).

2. Technical Aspects

The programs can handle selections with up to 3 stages, 2 different applicant populations, and 10 elementary predictors. The total number of applicants is limited to 100 for the SSOV and to 25 for the SSOF program.

3. Input SSOV Program

Observe that all input is in free format: Variables or vectores that have a name commencing with the letters I, J, K, L, M, N get INTEGER values (i.e., no decimal point). All other variables, vectors and matrices get FLOATING POINT values. The number of blanks between consecutive values on a line (record) can be freely chosen.

- # 1: **NP, NF, NA, NR(I) (with I = 1, NF)**
  - NP: the number of predictors
  - NF: the number of selection stages
  - NA: the number of applicants
  - (NR(I), I = 1, NF): vector with the number of applicants retained after the first, the second, and so on stage.
- # 2 and following: **WEIVAL(I,J)** (with I = 1, NF and J = 1, NP).
  The element WEIVAL(I,J) indicates the weight with which the Jth elementary predictor is used to compose the Ith stage-specific composite predictor.
- #3: **VALP(K)** (with K = 1, NP)
  The element VALP(K) indicates the validity of predictor K
- # 4 and following: **CORP(K, I)** (with K = 1, NP-1 and I = K+1, NP)
  The element CORP(K,I) corresponds to the correlation of predictor K and predictor I. Note that only the strict upper triangular correlation matrix should be given.

- # 5: **EFP(K)** (with K = 1, NP)
  The element EFP(K) represents the effect size of predictor K. Note that the effect sizes indicate the standardized mean difference between the majority applicant population and the minority population. If all applicants come from the same population, set EFP(K) equal to 0. for all predictors.

- # 6: **NG, PGMA, NCOMP**
  - NG: the number of different applicant populations; should be set equal to 2 even if the applicants come from a single population..
  - PGMA: the proportion of the majority applicant subpopulation within the total applicant population. If applicants come from one single population, set PGMA equal to 1.
  - NCOMP: NCOMP must be equated to 7.

- # 7: **EFC**
  EFC denotes the effect size of the job performance criterion. If all applicants come from the same population, set EFC equal to 0.

- # 8: **RERR, NPO**
  - RERR: relative error; set RERR equal to .0000001 or smaller (but not smaller than .000000001)
  - NPO: set NPO equal to 1000000

4. Input SSOF Program

- # 1: **NP, NF, NA, NMAJ, NR(I) (with I = 1, NF)**
  - NP: the number of predictors
  - NF: the number of selection stages
  - NA: the number of applicants
  - NMAJ: the number of majority applicants in the applicant pool
  - (NR(I), I = 1, NF): vector with the number of applicants retained after the first, the second, and so on stage.

- # 2 and following: **WEIVAL(I,J)** (with I = 1, NF and J = 1, NP).
  The element WEIVAL(I,J) indicates the weight with which the Jth elementary predictor is used to compose the Ith stage-specific composite predictor.

- #3: **VALP(K)** (with K = 1, NP)
  The element VALP(K) indicates the validity of predictor K

- # 4 and following: **CORP(K, I)** (with K = 2, NP and I = 1, K-1)
  The element CORP(K,I) corresponds to the correlation of predictor K and predictor I. Note that only the strict upper triangular correlation matrix should be given.

- # 5: **EFP(K)** (with K = 1, NP)
  The element EFP(K) represents the effect size of predictor K. Note that the effect sizes indicate the standardized mean difference between the majority applicant population and the minority population.

- # 6: **NG, NCOMP**
  - NG: the number of different applicant populations; should be set equal to 2.
  - NCOMP: NCOMP must be equated to 4.

- # 7: **EFC**
  EFC denotes the effect size of the job performance criterion

- # 8: **RERR, NPO**
  - RERR: relative error; set RERR equal to .0000001 or smaller (but not smaller than .000000001)
  - NPO: set NPO equal to 1000000

5. Sample Input File

Important: in preparing the input file, use a simple text editor such as Notepad, Wordpad or any other standard ASCII producing editor. DO NOT USE TEXT PROCESSING PROGRAMS SUCH AS MS-WORD or WORDPERFECT. Also, when saving the input file in Notepad, use the option "All Files" in the "Save as type" box. When saving in Wordpad, use the "Text Document-MS-DOS Format" option in the "Save as type" box, and be **aware that Wordpad has the nasty habit of adding the extension .txt to the file name that you specify**. Thus, with Wordpad, if you specify the name of the input file as "MINPUT", the file will in fact be saved as "MINPUT.TXT"; and this is the name that you have to use in the command to run the present programs. Here is a sample input file, for the `SSOV` program.

```
  5   2   10   6   2
  0.5   0.0   0.1   0.1   0.0
  0.0   0.1   0.0   0.0   0.1
 0.520   0.480   0.220 0.32   0.42
  0.310   0.030   0.370   0.020
```

```
 0.130  0.170 -0.020
 0.310  0.340
 0.250
0.72   0.32  0.06  .57  .04
 2  .8   7
 0.38
0.0000001  1000000
```

Next, a second input file for the SSOF program.

```
 5  2  10  8   6  2
 0.5  0.0  0.1  0.1  0.0
 0.0  0.1  0.0  0.0  0.1
0.520  0.480  0.220 0.32  0.42
 0.310  0.030  0.370  0.020
 0.130  0.170 -0.020
 0.310  0.340
 0.250
0.72   0.32  0.06  .57  .04
 2  4
 0.38
0.00000001  1000000
```

6. Running the Program

Suppose you copied the executable source of the program to the `d:ssel` directory on your machine. In that case, the input file must also be saved in the `d:ssel` directory. Next, to run the program, you have to open an MS-DOS Command window. The way to do this varies from one operating system (e.g., Windows 7) to the other, and you should use your local "HELP" button when in doubt about this feature.

In the MS-DOS Command window you type `d:`, followed by RETURN or ENTER, and your computer will return the `D:\>` command prompt. Next, you type `cd ssel` after the `D:\>` command prompt, again followed by RETURN or ENTER, and your computer will respond with the `D:\ssel>` command prompt. Now, you can execute the program by typing `caimsguz < minput > moutput` where "minput" is the name of the input file and "moutput" is the name of the output file. At the end of the execution, the PC will return the command prompt `D:\ssel>`. You can then inspect the output by editing the output file with either Notepad, Wordpad or any other simple editor program.

7. Sample Output

The first output corresponds to the first analysis, the second to the input for the SSOF program.

```
++++++++++++
+   SSOV   +
++++++++++++
```

```
Implementation of the VAP method (cf. De Corte, 2014) for the
computation of the EXPECTED CRITERION SCORE of the selected applicants,
the STANDARD DEVIATION of the AVERAGE CRITERION SCORE of the SELECTED,
and the AI RATIO for mixed, finite applicant pool multi-stage selections.

The program is at present limited to single, two and three
stage selection decisions, and to a maximum of 100 applicants.
The computations are based on results of
Norell, Arnason and Hugason, Biometrics, 1991.


Program written by Wilfried De Corte, Ghent University

The program uses routines from the Slatec library (see
http://www.geocities.com/Athens/Olympus/5564), the DCUHRE code from
Bernsten, Espelid and Genz (see
http://www.sci.wsu.edu/math/faculty/genz/software/dcuhre.f) and a couple
of algorithms from StatLib (see http://lib.stat.cmu.edu/apstat/)

PROBLEM SPECIFICATION

Number of stages in the selection:  2
Total number of applicants:     10
Number of retained applicants at each stage:      6      2
Number of Subpopulations:  2
Mixture Proportions of the Subpopulations:  0.80  0.20
Effect Sizes of the Predictors
   Subpopulation  1:  0.7200  0.3200  0.0600  0.5700  0.0400
   Subpopulation  2:  0.0000  0.0000  0.0000  0.0000  0.0000
Correlation of the Predictors with the Criterion:
             0.520   0.480   0.220   0.320   0.420
Correlation Matrix of the Predictors:
             1.000   0.310   0.030   0.370   0.020
             0.310   1.000   0.130   0.170  -0.020
             0.030   0.130   1.000   0.310   0.340
             0.370   0.170   0.310   1.000   0.250
             0.020  -0.020   0.340   0.250   1.000
Predictor weights in the stages
   0.5000   0.0000   0.1000   0.1000   0.0000
   0.0000   0.1000   0.0000   0.0000   0.1000
Effect Sizes of the Predictors in the Stages per Subpopulation
   Subpopulation  1:  0.7522  0.2571
   Subpopulation  2:  0.0000  0.0000
Criterion Effect Sizes per Subpopulation:  0.38  0.00
Correlation of the Stage-specific Predictors with the Criterion:
             0.558   0.643
Correlation Matrix of the Stage-specific Predictors:
             1.000   0.323
             0.323   1.000
```

```
Technical Input Parameters
   maxcls: 1000000    relative error: 0.000000100000

PROGRAM OUTPUT

Expected Criterion Score Selected:    1.1912

Expected selection rate of groups:    0.2212    0.1151
Expected AI ratio:    0.5203

Standard Deviation Average Criterion Score Selected:    0.6090

CPU TIME IN SECONDS         0.37


                         ++++++++++++
                         +   SSOF   +
                         ++++++++++++

Implementation of the FAP method (cf. De Corte, 2014) for the
computation of the EXPECTED CRITERION SCORE of the selected applicants,
the STANDARD DEVIATION of the AVERAGE CRITERION SCORE of the SELECTED,
and the AI RATIO for mixed, finite applicant pool multi-stage selections.
The computation is conditional given the fixed number of minority and
majority applicants

The program is at present limited to single, two and three
stage selection decisions, and to a maximum of 25 applicants.
The computations are based on results of,
Norell, Arnason and Hugason, Biometrics, 1991.


Program written by Wilfried De Corte, Ghent University

The program uses routines from the Slatec library (see
http://www.geocities.com/Athens/Olympus/5564), the DCUHRE code from
Bernsten, Espelid and Genz (see
http://www.sci.wsu.edu/math/faculty/genz/software/dcuhre.f) and a couple
of algorithms from StatLib (see http://lib.stat.cmu.edu/apstat/)

PROBLEM SPECIFICATION

Number of stages in the selection:  2
Total number of applicants:     10
Number of majority applicants:    8
Number of retained applicants at each stage:      6      2
Number of Subpopulations:  2
Effect Sizes of the Predictors
   Subpopulation  1:  0.72000  0.32000  0.06000  0.57000  0.04000
   Subpopulation  2:  0.00000  0.00000  0.00000  0.00000  0.00000
```

```
Correlation of the Predictors with the Criterion:
             0.520   0.480   0.220   0.320   0.420
Correlation Matrix of the Predictors:
             1.000   0.310   0.030   0.370   0.020
             0.310   1.000   0.130   0.170  -0.020
             0.030   0.130   1.000   0.310   0.340
             0.370   0.170   0.310   1.000   0.250
             0.020  -0.020   0.340   0.250   1.000
Predictor weights in the stages
   0.5000   0.0000   0.1000   0.1000   0.0000
   0.0000   0.1000   0.0000   0.0000   0.1000
Effect Sizes of the Predictors in the Stages per Subpopulation
   Subpopulation  1:  0.75225  0.25714
   Subpopulation  2:  0.00000  0.00000
Criterion Effect Sizes per Subpopulation:  0.38  0.00
Correlation of the Stage-specific Predictors with the Criterion:
             0.558   0.643
Correlation Matrix of the Stage-specific Predictors:
             1.000   0.323
             0.323   1.000
Technical Input Parameters
   maxcls: 1000000    relative error:   0.1E-07

PROGRAM OUTPUT

Expected Criterion Score Selected:    1.1928
Expected selection rate of groups:    0.2233    0.1069
Expected AI ratio:    0.4786
Standard Deviation Average Criterion Score     0.6078

CPU TIME IN SECONDS        14.26
```

8. Description of Output

The output is self-explanatory.

9. Dependencies and Acknowledgement

The present program is written in Fortran77. It was compiled to an executable code for WIN32 PCs (ie, Windows 7 or 8) with the GNU Fortran G77 compiler (cf. `http://www.geocities.com/Athens/Olympus/5564/`). The program uses routines from the SLATEC program library (cf. Fong et a., 1993; `http://www.geocities.com/Athens/Olympus/5564/`), a couple of algorithms from StatLib `http://lib.stat.cmu.edu/apstat/`), and some (adapted) code from Genz to evaluate multivariate normal probabilities (cf. http://www.math.wsu.edu/math/faculty/genz/homepage).

When the user reports results obtained by the present program, due reference should be made to De Corte (2014a) and De Corte (2014b).

10. References

De Corte, W. (2014a). SSOV and SSOF User's Guide.
Retrieved from `http://users.ugent.be/~wdecorte/software.html`

De Corte, W. (2014b). Predicting the outcomes of single and multistage selections when applicant pools are small and heterogeneous. Organization Research Methods, in press.